The ADAMACH Project

Adaptive And Meaning mACHines

Giuseppe Riccardi
Head of the Signals and Interactive Systems Lab
University of Trento, Italy

The ADAMACH Team

- Giuseppe Riccardi (PI)
- Postdocs
  - Alexei Ivanov (ASR)
  - Silvia Quarteroni (SLU)
  - Adam Sporka (HCI)
  - Sebastian Varges (DIALOG)

Human Interaction

Welcome to the Trentino Info Service! How May I Help You?

Uh hi, I need a hotel in Trento

Operator

Customer

Human-Human Conversation

Problem Solving Task

User Hi Good Morning

Operator Hi, How May I Help You?

U I am Roberta Sicconi calling from Cultural Affairs at City Hall.

Personal Identification

U I had made a request for a password change yesterday

O Ok do you have the request track id?

U Uhm No I cannot find

O Ok do you have the date of the request?

U Well that was yesterday

O...ok I think I can find it..I got it

Problem Statement Ticket Record Retrieval

O It's for a password reset.

U Right. The problem is that I changed the password when I first logged in..

Problem Resolution (USER)

O You were supposed to change first time you logged in. Now let's try together to log in

O can you tell me you RVS of your computer

U Well let me see. This is a new PC to me. Where can I find it?

O Usually the tag is right next to the base of the chassy next to the power switch. It reads "inventario settore informatico".

U Inventario..Settore... Informatico. Got it 123456

Problem Resolution (PART I) OPERATOR asks help to the USER to connect to his PC

O yes that is right. Now, you see I'm writing the old login..now you type in the new login. It should be at least 6 characters...

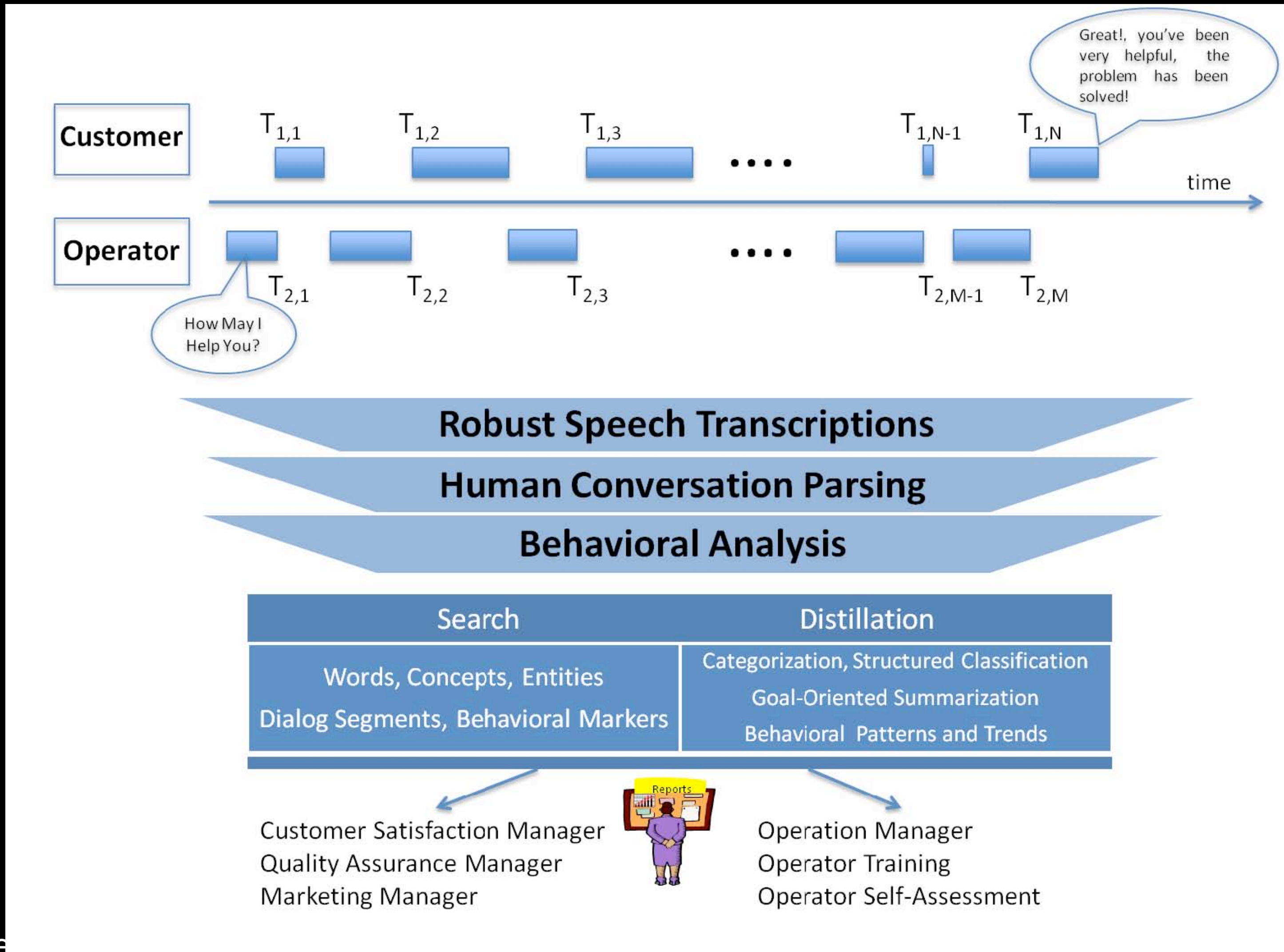
U Ok let me write that down one moment

....

Problem Resolution (PART II) OPERATOR and USER work together to fix the problem

Interactive Systems

Analytics Technology

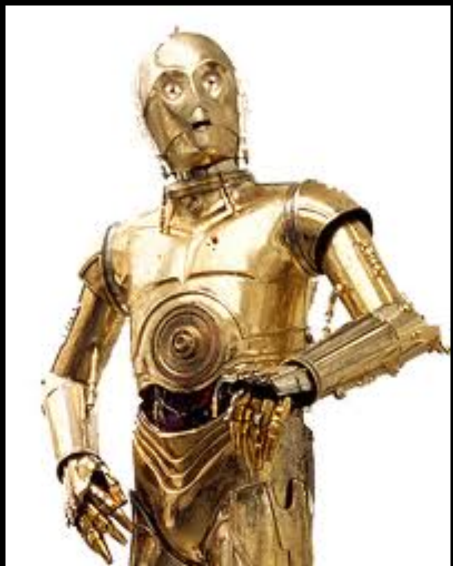


Interactive Systems

Conversational Agent Technology

Welcome to the Trentino Info Service! How May I Help You?

Uh hi, I need a hotel in Trento



Operator

Customer

Outline

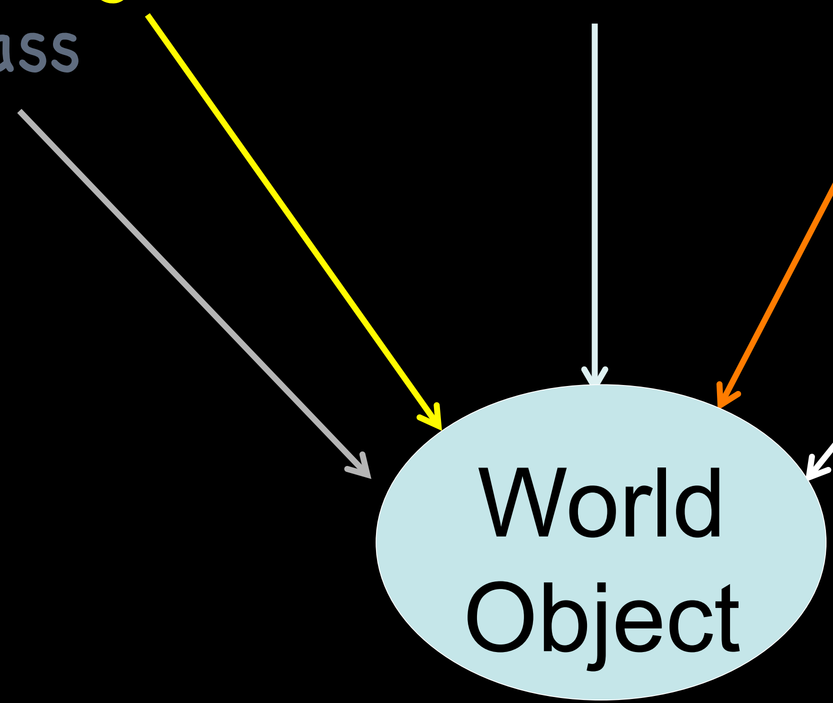
- Understand words/concepts
  - Linguistic vs Knowledge Structure ?
- Spoken Language Understanding
  - Robust Parsing models
- Adaptive Dialog Models
  - Rule-based vs Statistical Models
- Personable conversational agents

Spoken Language Understanding

- Core component of Spoken Dialog Systems
- "Voice search" applications
  - Smartphones
  - Short speech cycle (video)
- Grammar-based vs Statistical Models
- Understand words/concepts
  - Signal – to – Symbol Mapping
  - Traditionally grounding is done over the words

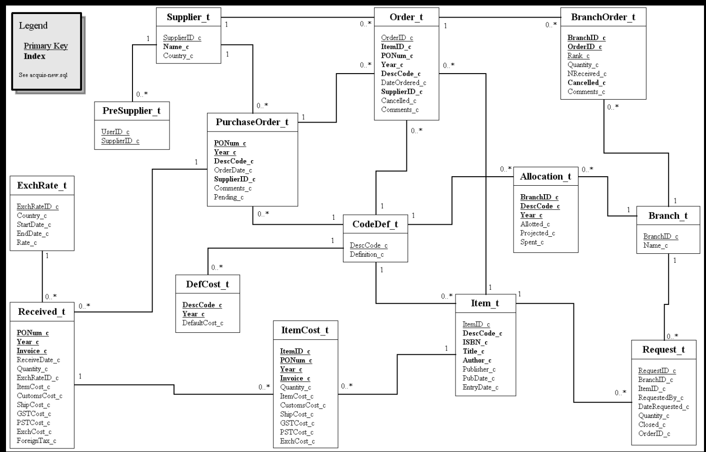
Language Understanding

Find the best flight from New York to Paris tomorrow business class



World Object

Databases, Ontologies



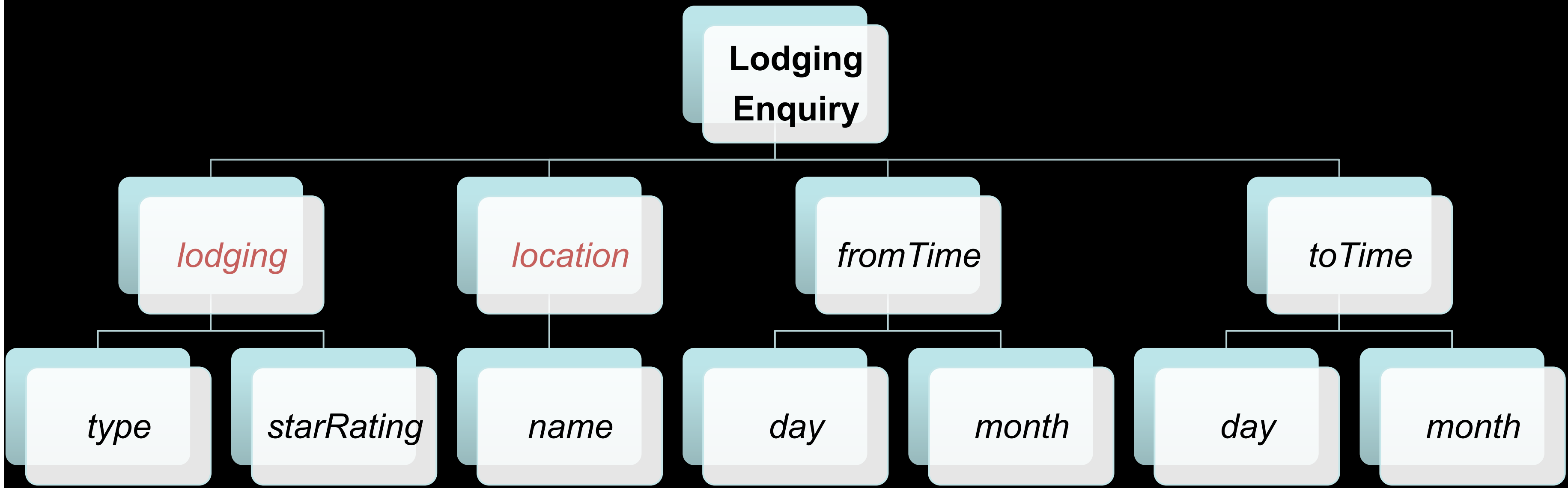
Semantic Web is not AI
(1998)

The concept of machine-understandable documents does not imply some magical artificial intelligence which allows machines to comprehend human mumblings..... **Instead of asking machines to understand people's language, it involves asking people to make the extra effort.**

(T.B.Lee, 1998)

Domain Ontology

Tourist Domain: LodgingEnquiry



Language Understanding

Find the best flight from New York to Paris tomorrow

business class

USER
CONSTRAINTS

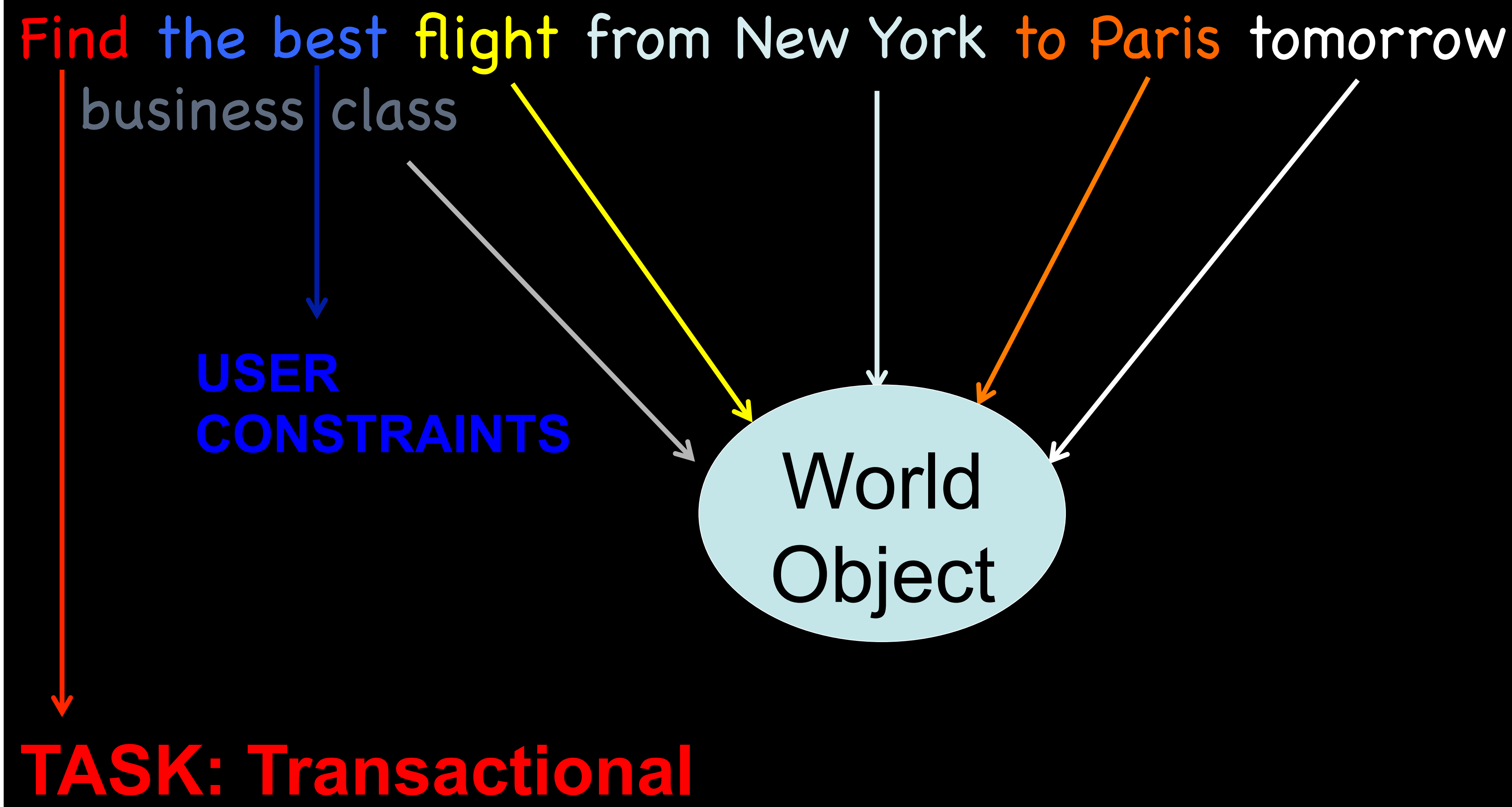
Language Understanding

Find the best flight from New York to Paris tomorrow

business class

TASK: Informational

Language Understanding



Spoken Language Understanding

- **What the USER says:**
  "Find the best flight from New York to Paris tomorrow business class"
- **What the Machine believes user said:**
  " Find the bass flight from Newark to Paris tomorrow business class"
- **What the Machine believes user meant:**
  - @action=Request-Reservation (0.9)
  - @origin=Newark (0.5)
  - @time-departure=Tuesday (0.7)
  - @destination=Paris (0.8)

Speech Recognition

Parsing

SLU Models

- Goal: Observations X must be assigned labels from Y
  - X=word sequence, Y=concept sequence
- Two main approaches:

GENERATIVE	DISCRIMINATIVE
P(X,Y)=P(X|Y)*P(Y)	*f*: X → Y as P(Y|X)
• Hidden Vector State model *(He&Young 2005)* • Statistical Machine Translation *(Hahn et Al. 2008)* • Stochastic Finite State Transducers *(Raymond et Al. 2006)*	• Support Vector Machines *(Vapnik 1998)* *(Raymond&Riccardi 2007)* • Log-Linear models (ME,CRF) *(Bender et Al. 2003)* *(Hahn et Al. 2008)* *(Lafferty et Al. 2001)*

Discriminative Reranking

- Discriminative Reranking Models (DRMs)
  - Combines the best of both approaches
  - Ourperforms best segmentation/labeling model (CRF)
  - Extendable to other parsing models (grammar-based)

Discriminative Reranking Models for Spoken Language Understanding,
M. Dinarelli, A. Moschitti and G. Riccardi, IEEE Transactions SLP to appear 2011

Knowledge Representation vs Semantic Representation

- Traditionally ad-hoc domain concept representations are used
- Poor coverage and portability across domains, systems and applications
- Semantic representation
  - Lexicalized Resource
  - Large coverage (domain and language)
  - Interface with world objects

FrameNet Semantics

- **Semantic frame**
  - E.g. ***REQUEST***

    Definition*: In this frame a Speaker asks an Addressee for something, or to carry out some action.*
- **Lexical Unit (LU):**
  - E.g. in ***REQUEST***:

    ask, beg, command, demand, implore, order, petition, request, urge
- **Frame Element**:
  - E.g. in ***REQUEST*** :

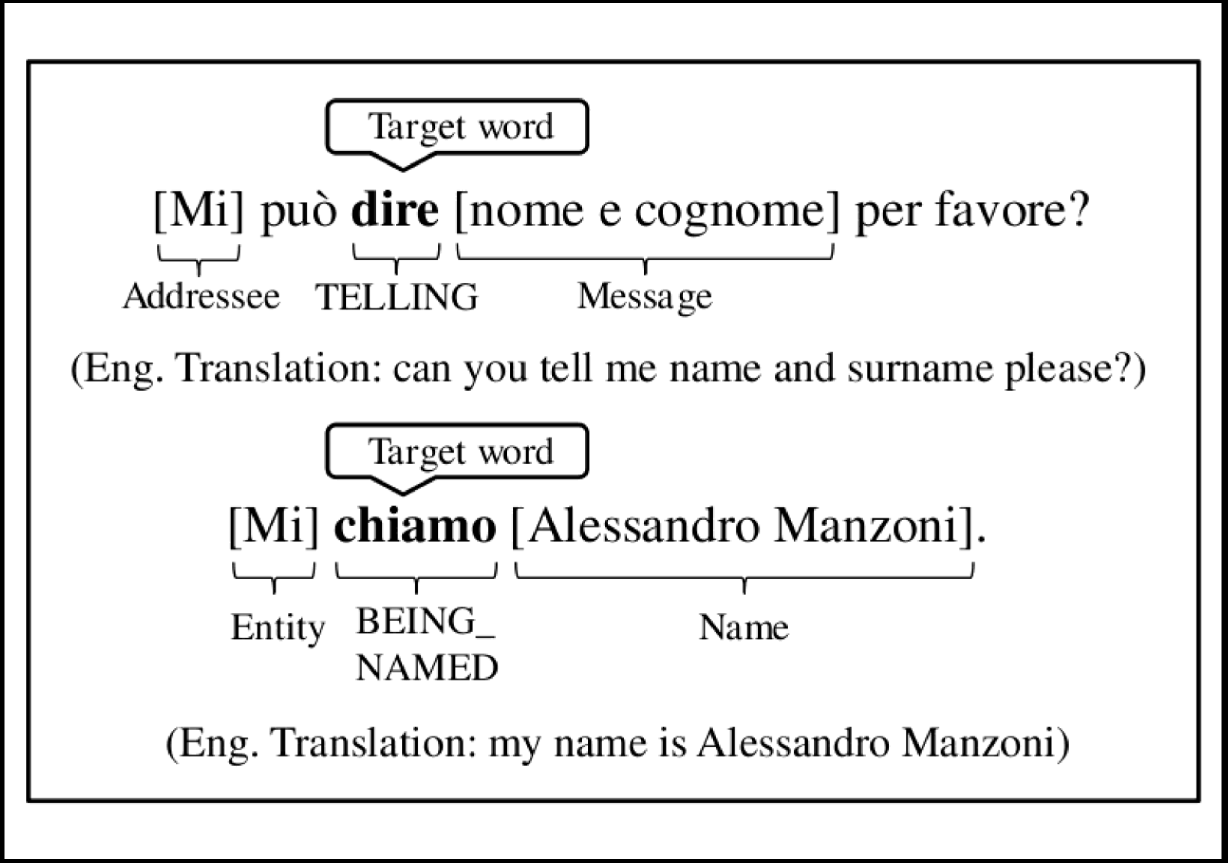
    Core*: Speaker, Addressee, Topic, Message, Medium*

    Non core*: Beneficiary, Manner, Means, Time*

In fact [I]$_{\text{Addressee}}$ was ASKED [to chair the meeting]$_{\text{Message}}$

[Tong]$_{\text{Speaker}}$ ORDERED [the pilot]$_{\text{Addressee}}$ [to circle Ho Chi Minh City]$_{\text{Message}}$

Annotation of Spoken Dialogs



"dire", "chiamo" - **target words**, which recall a **Semantic Frame**

Annotating Spoken Dialogs: from Speech Segments to Dialog Acts and Frame Semantics
M. Dinarelli et al. , EACL Workshop on Semantic Representation of Dialogue, 2009

Frame-based Parser

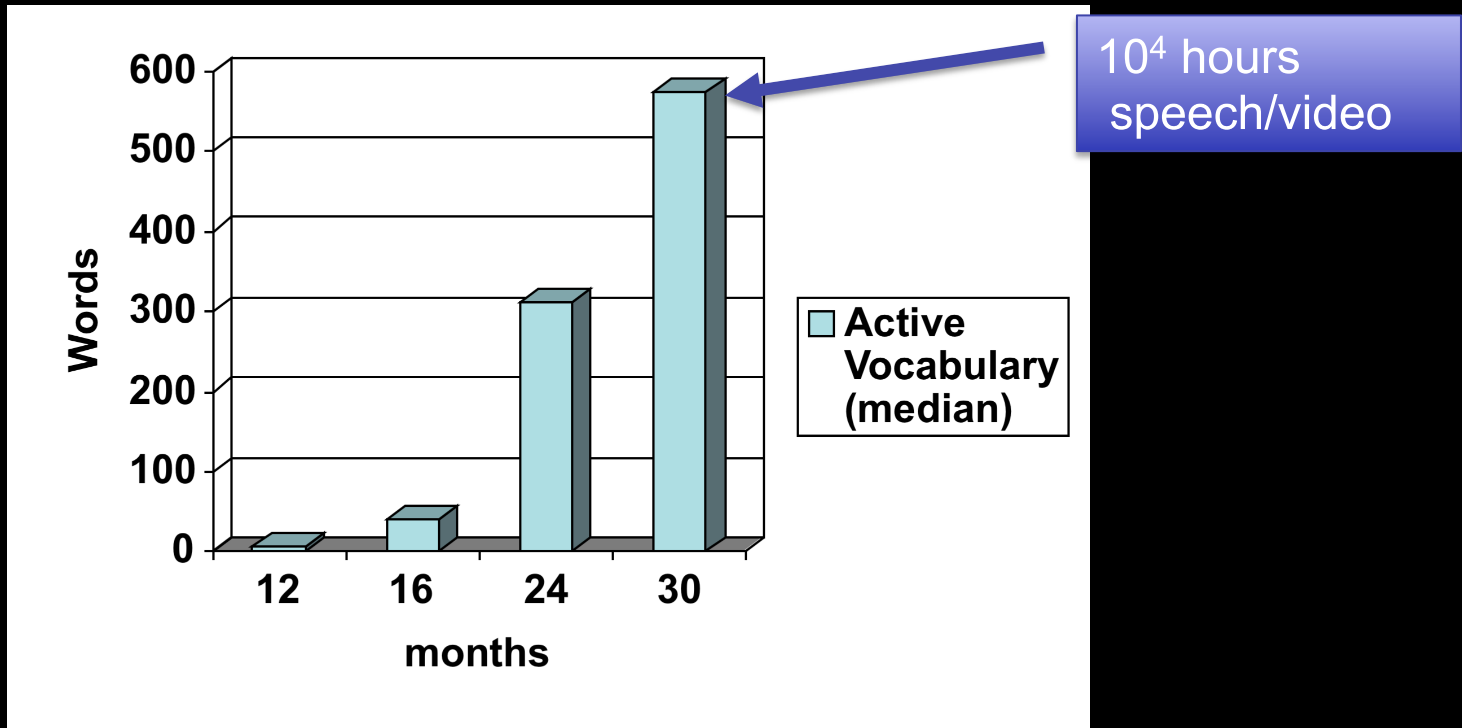
- Plain text sentence *(syntax omitted)*:
  Ralemberg said he already had a buyer for the wine.
- Target Word Selection (dictionary keyword: *buyer*)
  Ralemberg said he already had a buyer for the wine.
- Frame Disambiguation:
  Selected Frame: Commerce_Scenario
- Argument Boundary Detection:
  Ralemberg said [he] already had a [buyer] [for the wine].
- Argument Role Classification:
  *Ralemberg said [he]*SELLER *already had a [buyer]*BUYER *[for the wine]*GOODS.

B. Coppola, A. Moschitti and G. Riccardi
NAACL 2009

Grounding Meaning Directly into Speech Features

Acoustic Correlates of Meaning

Infants' Language Acquisition



Estimate of infants' productive vocabulary size

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D., & Pethick, S.J. (1994) "*Variability in early communicative development*", Monographs of the Society for Research in Child Development, 59 (5 serial no. 242)

Grounding Meaning into non-lexical Speech Features

- Parsing **of meaning structures** is traditionally carried **over** the **word hypotheses** generated by the ASR.
- How to **discover meaning components from** direct measurements of **acoustic features**?
- Such features may be more robust and complementary to lexical features.

Acoustic Features

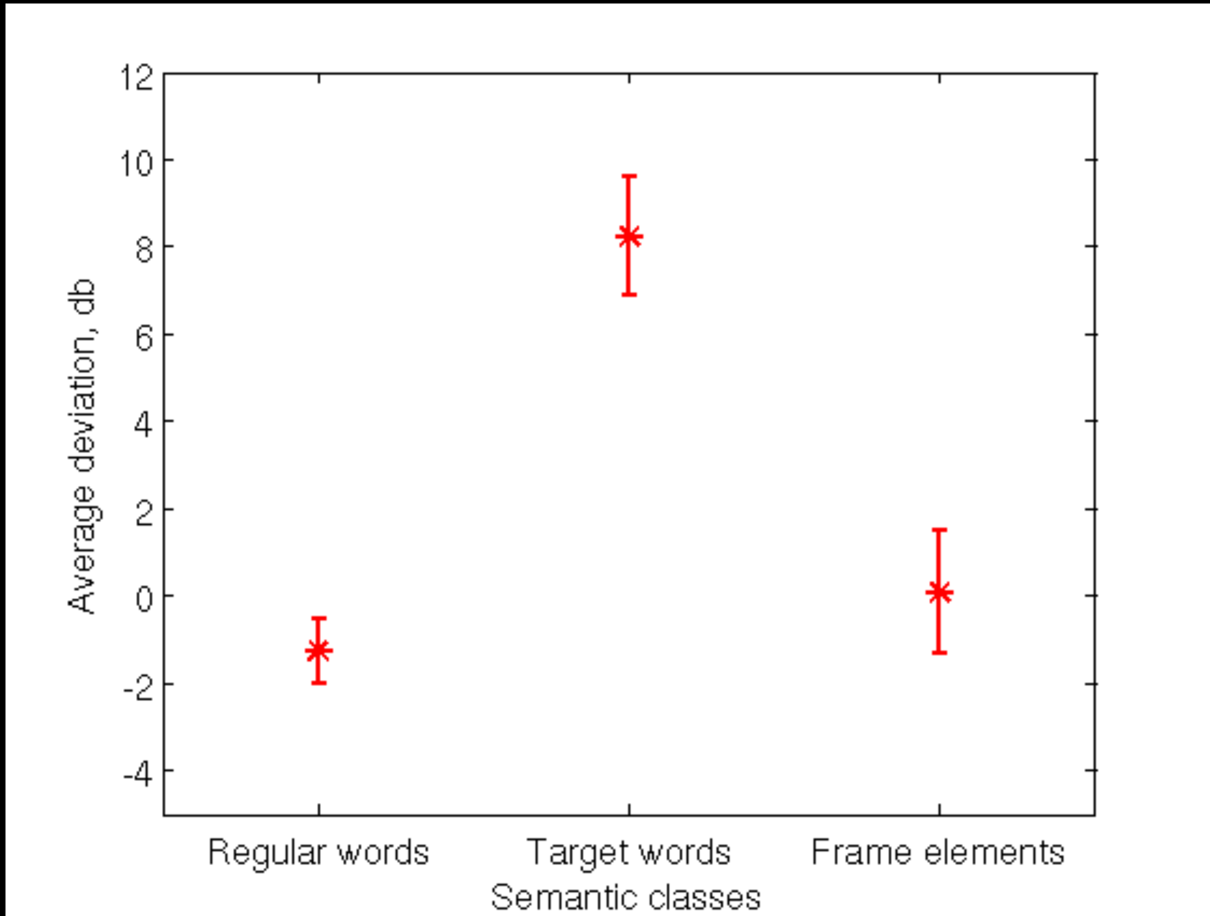
- **Pitch, voiced interval duration, formant trajectories, total intensity(I_{tot}), harmonicity (I_{hnr}).**
- **Intensity and Harmonicity were combined** to obtain an intensity of harmonic speech component I_{harm}.

$$I_{harm} = I_{tot} + I_{hnr} - 10\log_{10}(10^{I_{hnr}/10} + 1)$$

- I_{harm} **reflects intensity of phonation** (voicing) **rather then** sound production by **friction** or **obstruction**.

Acoustic-semantic correlates

Harmonic Intensity



Prediction of Target Words

Lexical & Acoustic Features

Classifier	Prec.	Recall	F1
Best Linguistic	0.782	0.841	0.810
Best Acoustic	0.247	0.774	0.375
Oracle Combination	0.935	0.913	0.924
Baseline Linguistic	0.759	0.648	0.699
Oracle Comb. (+ best acoustic)	0.926	0.811	0.865

Acoustic Correlates of Meaning Structure in Conversational Speech

A. Ivanov, G. Riccardi, S. Ghosh, S. Tonelli and E.A. Stepanov, Interspeech 2010

Outline

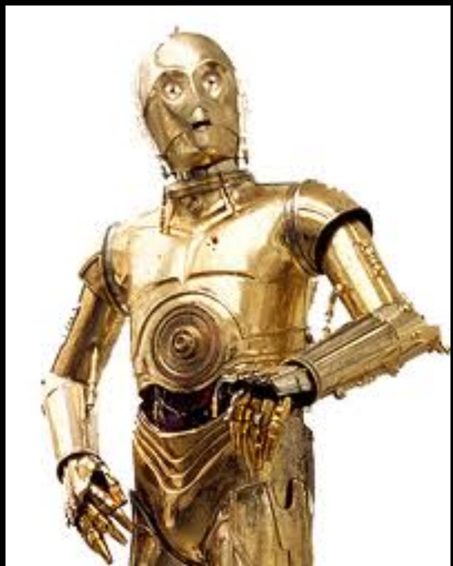
- Understand words/concepts
  - Linguistic vs Knowledge Structure ?
- Spoken Language Understanding
  - Robust Parsing models
- Adaptive Dialog Models
  - Rule-based vs Statistical Models
- Personable conversational agents

Interactive Systems

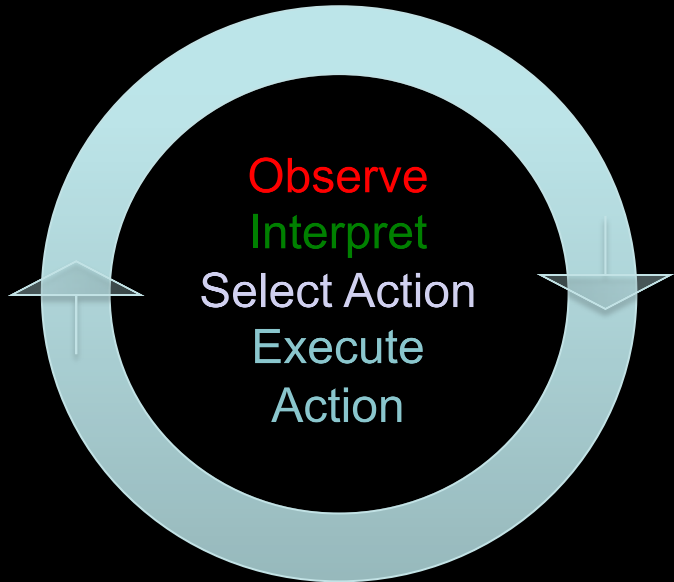
Conversational Agent Technology

Welcome to the Trentino Info Service! How May I Help You?

Uh hi, I need a hotel in Trento



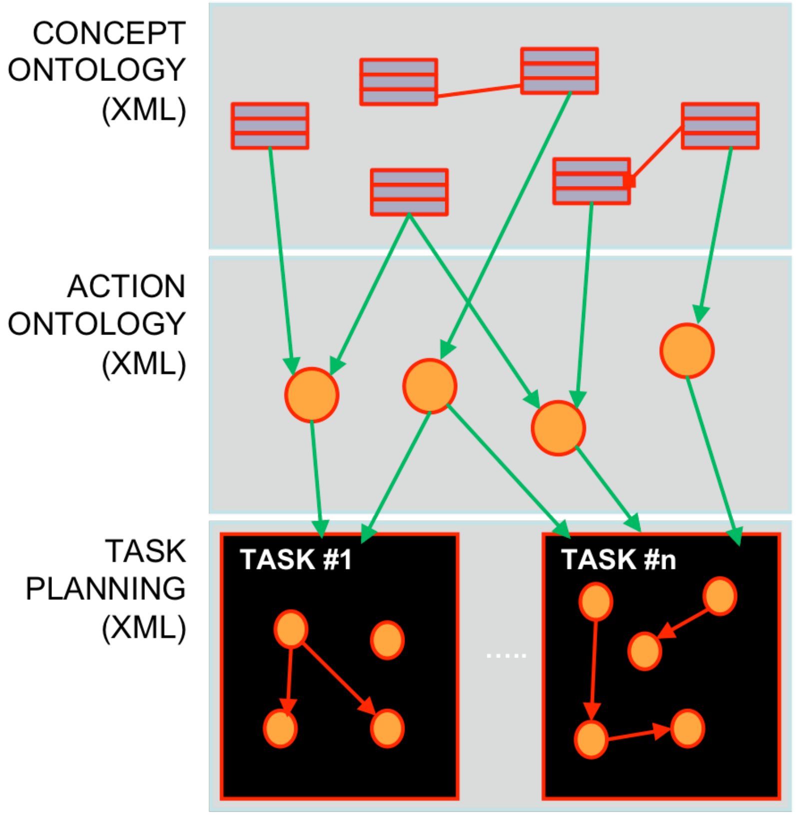
Operator



Customer

Dialog Models (DM)

Rule-Based Systems



1) Domain Representation

2) Task Representation

3) Task Execution

are hand-coded

Example Interpretation Rule

"*Match user provided digits if system has asked for student ID in last turn, and either accept or verify the digits as a student ID, depending on confidence*"

```
(defrule match-answer-student-id
    (last-system-move (move question student-id)
                      (expect answer student-id))
    (last-user-turn   (interp-attr digits)  (interp-val ?digits)
                      (confidence ?confidence))
    =>
    (if (>= ?confidence ?*threshold-conf-student-id*)
        then (assert (application-parameter
                           (parameter student-id) (value ?digits)))
        else (assert (verification-required
                           (parameter student-id) (value ?digits)))))
```

Pros/Cons

- Interaction Control
  - Human-coded Strategies
  - Direct (Human Interpretable Rules)
  - Heuristics-driven (e.g. Business Rules)
- Human-Free Control
  - Automatic Learning of Strategies (e.g. unseen events, observations)
  - Task Complexity Management
  - Multimodal Language, Observations of the world state

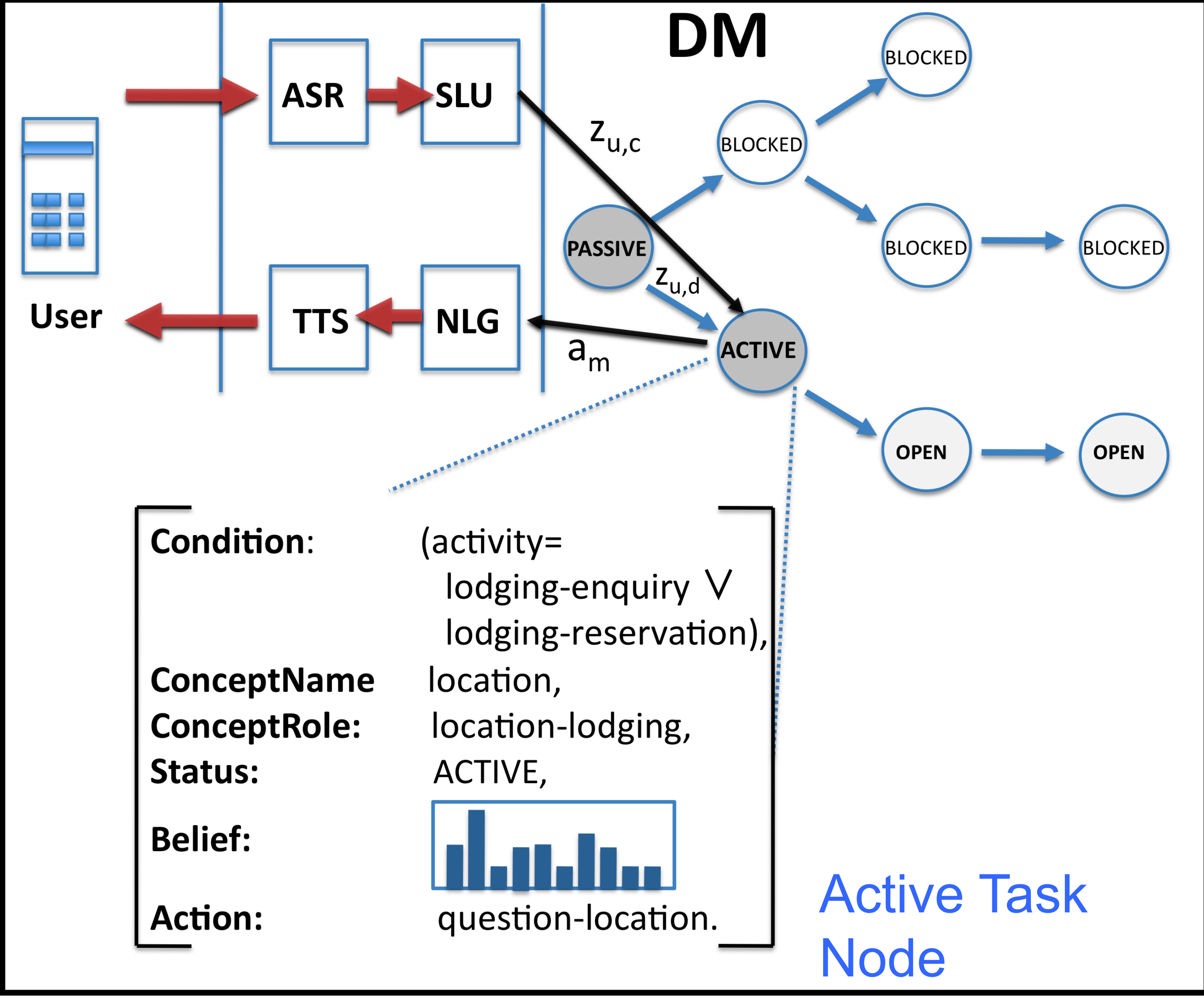
Reinforcement Learning

- Learning from interaction of agent with its environment
- Uncertainty about the environment:
  - exact planning not possible in general
  - instead simulations are used (`trial-and-error')
- Reward
  Defines the cost structure of the interaction (from system and user perspective)

Markov Decision Processes

- Statistical Modeling of Human-Machine Interaction
- MDPs vs Partially Observable MDPs
- Uncertainty in the User Input semantic interpretation (MDP)
- Uncertainty in the User State (POMDP)
- Autonomous Learning of dialog strategies
- Reward-driven learning

Hybrid POMDP-DM



Effect Clarification Strategies

	RULE	
	first	final
activity	78	74
loction	64	74
starrating	67	70
month	85	89
day	70	76
ALL	**74**	**78**
σ=	(0.02)	(0.02)

Metric:
Precision of first and final mentions of concept value measures effectiveness of clarification strategies

- RULE: Rule-based dialog manager

Effect Clarification Strategies

	RULE			POMDP-Ia			POMDP-Ib			POMDP-IIb		
	first	final	$\delta\%$	first	final	$\delta\%$	first	final	$\delta\%$	first	final	$\delta\%$
activity	78	74	-4.1	83	88	5.0	83	96	15.7	84*	84*	0.0
loction	64	74	15.8	69	73	6.3	54	69	28.0	66	76	14.3
starrating	67	70	3.4	90	97	7.7	87	96	10.0	94	96	2.6
month	85	89	4.3	76	86	12.7	76	83	9.0	92	93	1.6
day	70	76	8.3	61	76	25.3	74	82	10.0	76	90	18.3
ALL	**74**	**78**	**5.2**	**74**	**83**	**12.1**	**74**	**84**	**13.3**	**82**	**88**	**7.4**
$\sigma=$	(0.02)	(0.02)	(2.11)	(0.03)	(0.03)	(2.19)	(0.03)	(0.03)	(2.50)	(0.02)	(0.02)	(2.09)

- RULE: Rule-based dialog manager
- POMDP-Ia: POMDP-DM
- POMDP-Ib: POMDP-DM with advanced SLU
- POMDP-IIb: Confidence POMDP-DM w/ adv. SLU

POMDP Concept Policies for Hybrid Dialog Management
S. Varges, G. Riccardi, S. Quarteroni and A. Ivanov, ICASSP 2011

Task completion and length metrics

	Lodging Task		Event Enquiry		ALL
	TCR	#turns	TCR	#turns	TCR
RULE	70.3% (26/37)	13.0 (σ=3.5)	66.7% (28/42)	8.7 (σ=2.5)	68.4% (54/79)
POMDP-Ia	79.0% (45/57)	22.0 (σ=5.8)	84.3% (27/32)	14.4 (σ=4.3)	80.9% (72/89)
POMDP-Ib	91.4% (74/81)	19.8 (σ=4.1)	94.2% (65/69)	13.3 (σ=2.9)	92.7% (139/150)
POMDP-IIb	88.8% (87/98)	21.7 (σ=5.5)	86.5% (45/52)	14.0 (σ=5.2)	88.0% (132/150)

Trade-off between length and precision/success: POMDP is optimized to improve precision

Exploration vs Exploitation

- Current dialog systems do not explore, rather exploit hardwired and expensive heuristic strategies.
- Conversational Agent needs to find trade-off between exploration and exploitation
- No separation between training and testing:
  - most natural for RL and in 'real world',
  - continues to learn/adapt (learning rate)

DEMO

http://youtu.be/3QY-IkIvOHY

POMDP Concept Policies for Hybrid Dialog Management
S. Varges, G. Riccardi, S. Quarteroni and A. Ivanov, ICASSP 2011

Interaction Corpora are Expensive (TIME, NOISE, $)

Machine-to-Machine Interaction

Welcome to the Trentino Info Service! How may I help you?

Operator

Uh hi, I need a hotel in Trento

Customer

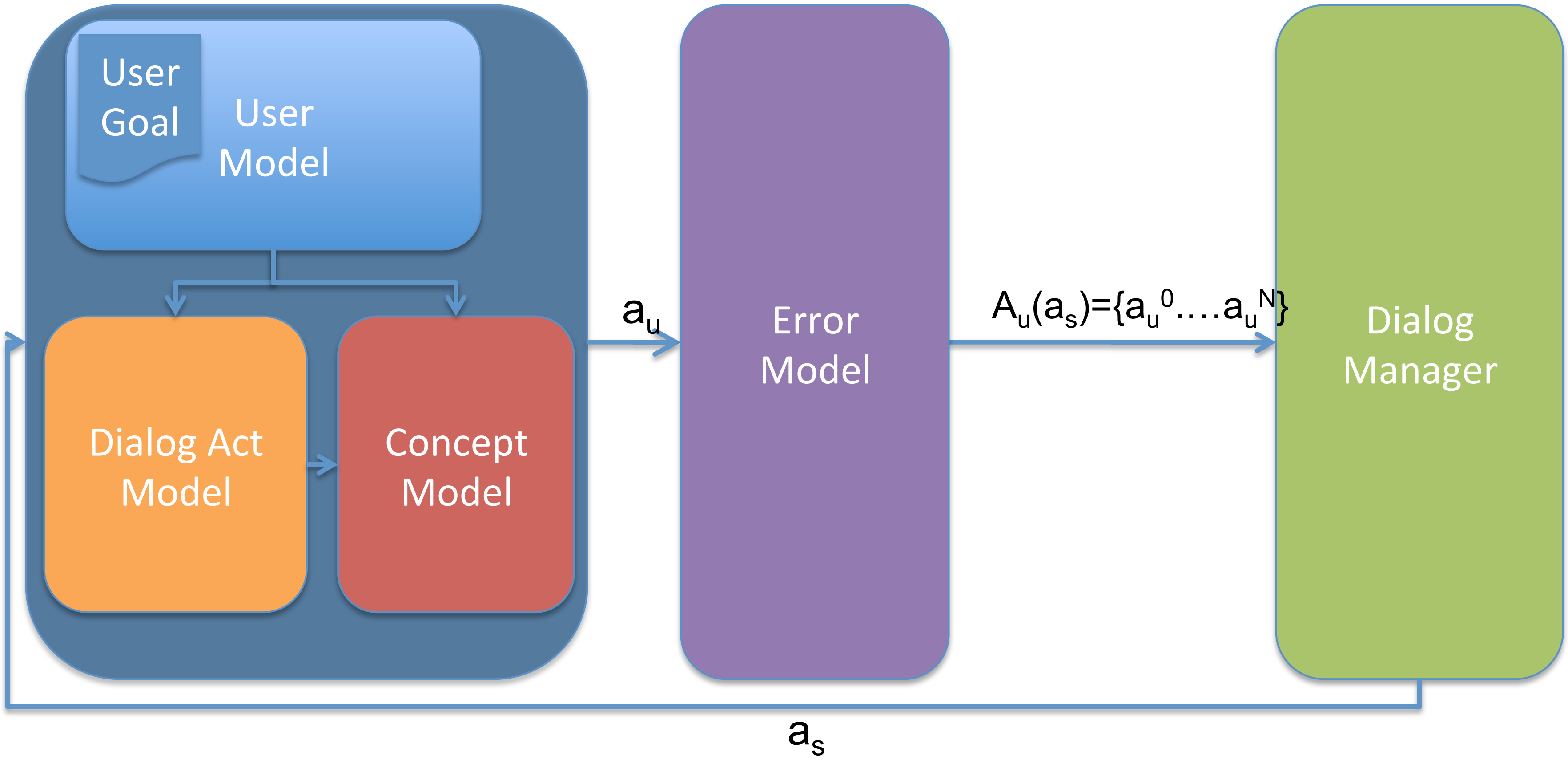
Statistical Interaction Simulation

- Train and evaluate the performance of Dialog Managers over infinite amount of interactions
- Experimental validation
- Data-driven simulator trained from real dialogs
  - Representation Level (words, user intentions, etc.)
  - Error Models (ASR, SLU, etc.)
  - User Behavior

Example DM – Simulator dialog

- DM: [Greet(); Offer()]
- SIM: [Info-request(activity = *EventEnquiry*; type = *expo*)]
- DM: [Info-request(location)]
- SIM: [Answer(location = *Vela*)]
- DM: [Info-request(month)]
- SIM: [Answer(month = *Nov*)]
- DM: [Clarif-request(month = *Nov*)]
- SIM: [Yes-answer()]
- …

Dialog Simulation Architecture



M. Gonzalez M., S. Quarteroni, G. Riccardi and S. Varges,
"Cooperative User Models in Statistical Dialog Simulators" SIGDial 2010

DEMO

http://youtu.be/eYvRWSa7zSY

Outline

- Understand words/concepts
  - Linguistic vs Knowledge Structure ?
- Spoken Language Understanding
  - Robust Parsing models
- Adaptive Dialog Models
  - Rule-based vs Statistical Models
- Personable Conversational agents

Motivations

- To identify the needs/preferences of the users
  - Should we make machines interact more human-like?
- Be aware of the speaker state
  - Emotion Recognition (late '90s, early 2000)
  - Personality Recognition (Mairesse et al. 2007, Polzehl et al. 2010, Ivanov et al., 2011)

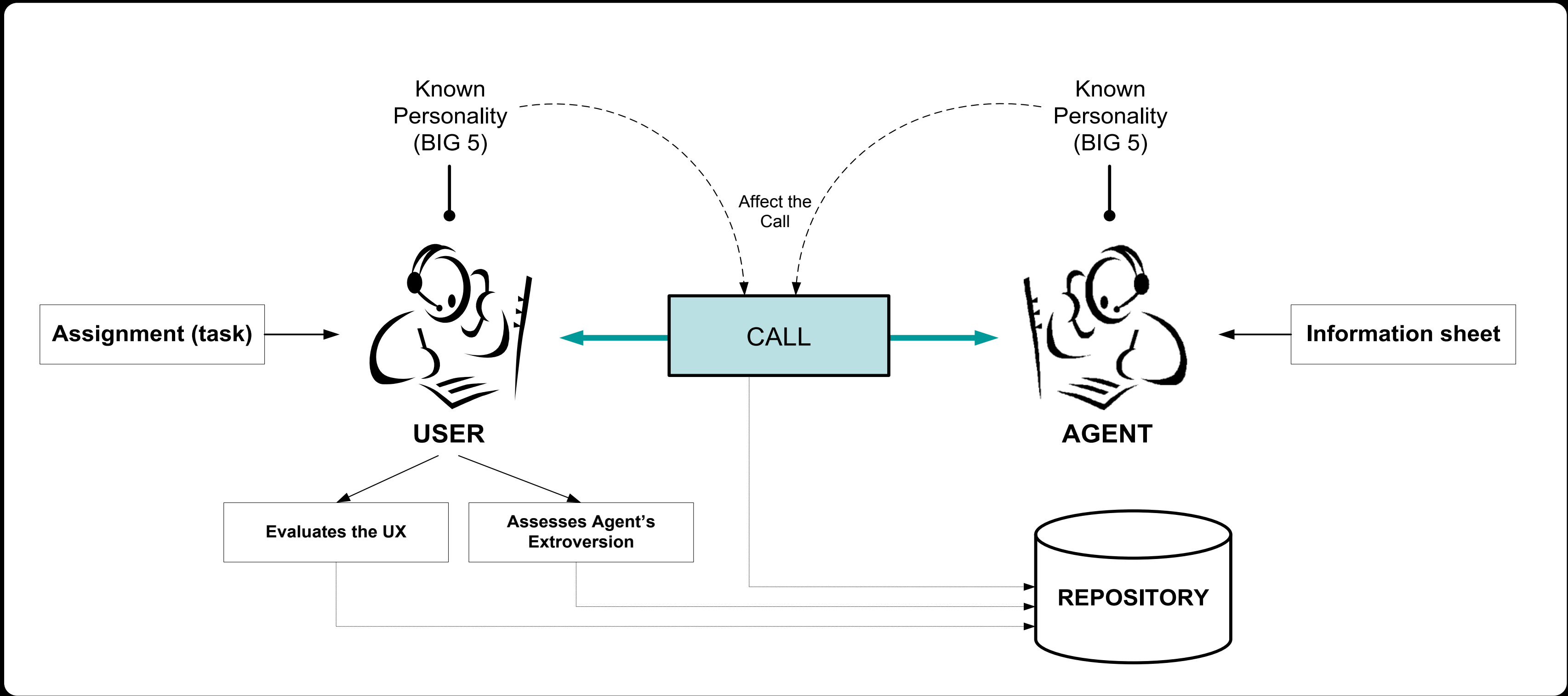
Personable Agents

- Role of *Personality* in communicating agents
- Personality modeling and generation supports
  - social layer of communication (personality matching)
  - dialog strategies (e.g. content generation & selection)
  - user modeling (e.g. emotion recognition/ synthesis)

Conversational Agents (1)

- Current models of conversational agents
  - Example
- Personality modeling and generation supports
  - social layer of communication (personality matching)
  - user modeling (e.g. emotion recognition)
  - dialog strategies (e.g. content generation selection)
- Examples
  - Extrovert / Introvert
  - Introvert / Introvert

PERSIA CORPUS



Recognition of Personality Traits from Human Spoken Conversations
A. Ivanov, G. Riccardi, A. Sporka and J. Franc, to appear Interspeech 2011

Data Collection

- 24 participants: 12 Users, 12 Agents
- **Personality traits**
  - BIG 5 personality traits of the interlocutors
  - Agent's extroversion, as perceived by the User (via post-task questionnaire)
- **Evaluation** (1 = lowest score, 7 = highest score)
  - User Experience variables according to ISO 9241-11: Effectiveness, Efficiency, Satisfaction

Speaker Personality Classification

Big-Five Personality Traits:

Openness to experience: A preference to a varying experience, an appreciation for art, emotion, adventure, etc.

Conscientiousness: A tendency to have a planned behavior (as opposed to spontaneous responses), a manifestation of self-discipline.

Extroversion: ``Energetic'' behavior, an outgoing attitude, seeking the company of others.

Agreeableness: Compassion and cooperativeness (as opposed to suspicion)

Neuroticism: A tendency to ``mood swings'', a tendency to negative emotions such as anger or vulnerability.

Personality Classifier

Paralinguistic features are extracted from the **whole dialog sides** (composition of all turns of a speaker in the dialog)

Feature Extraction:

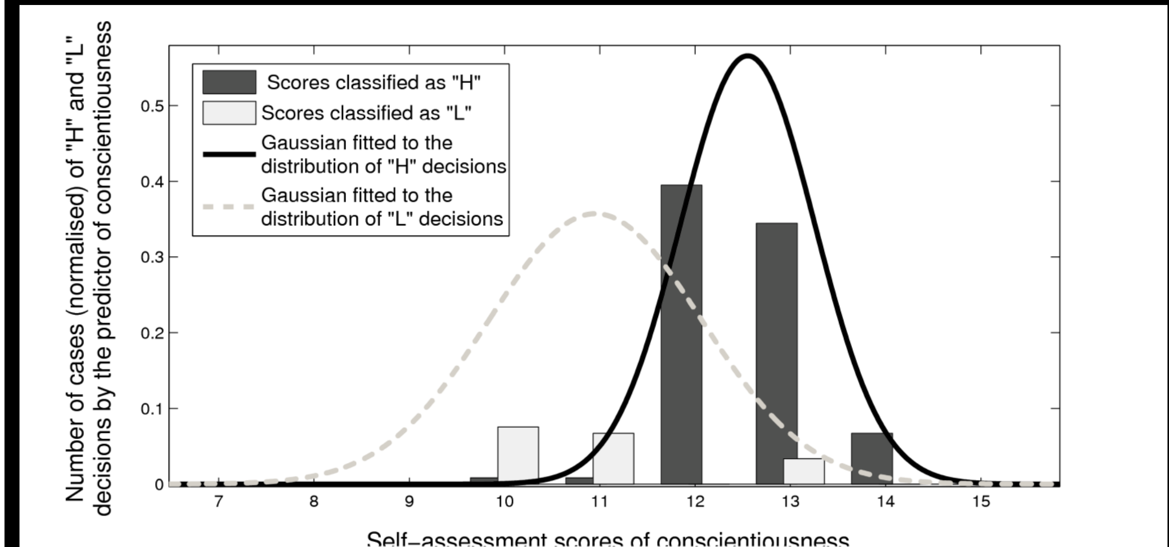
Based on OpenEar (**http://openart.sourceforge.net/**)

Classifier is based on Boostexter

http://www.cs.princeton.edu/~schapire/
boostexter.html

Classification Results

Personality Trait	CORR	Acc. %	Chance %	p-value
Openness	48	40.34	52.97	0.9962
Conscientiousness	**113**	**94.96**	**73.17**	$\mathbf{9.8 \cdot 10^{-11}}$
Extroversion	**75**	**63.03**	**50.00**	$\mathbf{1.6 \cdot 10^{-3}}$
Agreeableness	67	56.30	54.83	0.3401
Neuroticism	39	32.77	50.00	0.9999



Measurements were done in ***LOSO*** fashion

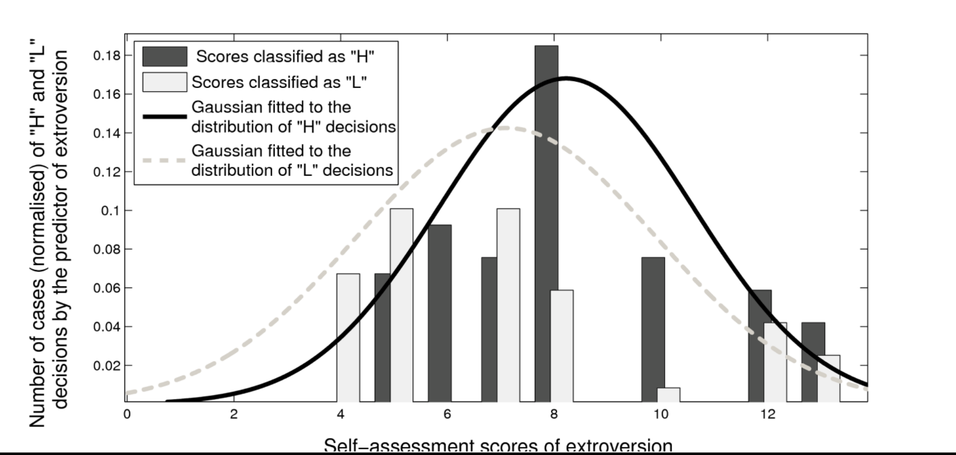
Conscientiousness is the **most reliably detectable** personality trait

Result is **statistically significant**

Classification Results

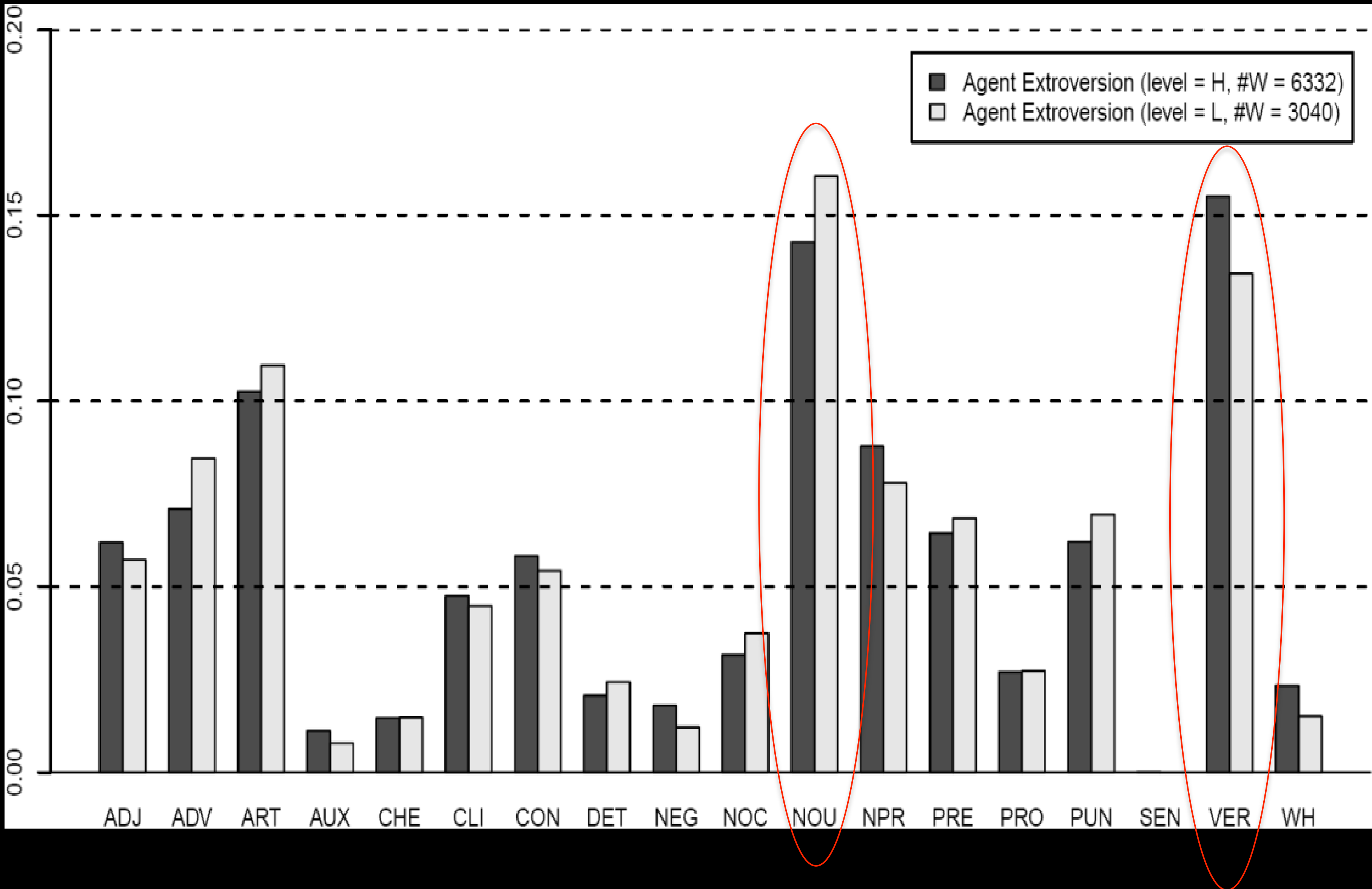
Extroversion

RM scores	CORR	Total	Acc. %	Chance %	p-value
6	75	108	69.44	50.43	$2.0 \cdot 10^{-5}$
6, 7	63	87	72.41	56.35	$6.8 \cdot 10^{-4}$



- **Extroversion** detection is also **above the chance performance**
- This is a **statistically significant** result
- However the **overlap between assigned labels is much greater** then with conscientiousness (see the figure above)
- If the **intermediate cases** (self-assessment scores 6 & 7) **are omitted** the result is **much better**
- The system **is good in detecting** the cases of extreme **extroversion and introversion**

Personality Affects Language



Conclusion

- **Communicative bottlenecks**
  - Recognition vs Understanding (e.g. 10^{6} ASR dictionary vs SLU 10^{2} concepts)
  - Multimodal Multisensorial Language Understanding/ Generation
- **Adaptive Machines**
  - Learning Systems (active learning -> active systems)
  - Context-aware communication (device, physical space, social roles)
  - Personal Agents

For More Information check:

www.sisl.disi.unitn.it